



King Saud University
Journal of the Saudi Heart Association

www.ksu.edu.sa
www.sha.org.sa
www.sciencedirect.com



EDITORIAL

The current state of risk stratification and EuroSCORE in cardiac surgery

Readers of the Saudi Heart Journal will be familiar with the European System for Cardiac Operative Risk Evaluation (EuroSCORE) as the world's most widely used cardiac surgical risk model (Nashef et al., 1999). Recent papers have cast doubt on the validity of this model, now more than 10 years old, for risk assessment in the second decade of the third millennium. It is worthwhile to pause and reflect a little on where we are now in relation to risk assessment.

1. Why use a risk model?

There are many reasons for using a risk model, but they can be broadly classified under two subheadings: assessing the risk for the patient and evaluating the quality of care provided by the institution.

Assessing risk

All decisions regarding health care depend on weighing the benefit against the likely risk, and this is true when a surgeon considers offering a treatment to a patient and when the patient gives informed consent to receiving the treatment. Indeed, when operation is being contemplated on prognostic grounds alone, in an asymptomatic patient, then evaluating the risk of surgery becomes of paramount importance: we must not offer an operation to an asymptomatic patient if surgery carries a greater risk than conservative treatment.

Evaluating care

A risk model offers a standard expected outcome against which the actual outcome can be measured. If the risk model says that your predicted mortality should be 5% for a certain group of patients with a particular risk profile, and your actual mortality is 1%, then you are doing well. Dividing actual mortality by your predicted mortality is termed the “risk-adjusted mortality ratio” or RAMR. This

can be given with a 95% confidence interval and is probably the most useful single measure of the performance of a cardiac surgical unit.

2. What makes a good risk model?

The validation of a risk model depends on the assessment of two features: calibration and discrimination. Calibration is the accuracy of the model for predicting risk in a group of patients, in other words, if the model says that mortality in a thousand patients is likely to be 5%, and actual mortality is 5% or close to 5%, then the model is well calibrated. Discrimination refers to the model's ability to distinguish between low-risk and high-risk patients. In other words, if most of the deaths occur in patients that the model correctly identifies as high-risk, the model has good discrimination, but if most deaths occur in patients that the model actually identifies as low-risk, there is poor discrimination. We measure discrimination using a statistic called the ‘area under the receiver operating characteristic (ROC) curve’, sometimes also called the *c*-statistic or the *c*-index. If the area under the ROC curve is 0.5, the model does not discriminate at all. Good discrimination begins at 0.7 and rarely exceeds 0.85. If the area under the ROC is 1.0, the model is no longer a risk model but a crystal ball which forecasts the future (an impossible task).

It is possible for a risk model to have good calibration but poor discrimination, and vice versa. Discrimination is more important than calibration. A model can be recalibrated or adjusted as practice improves, but if the model is built on the wrong risk factors, its discrimination cannot be improved.

3. Making electricity in Chicago

EuroSCORE was derived from data on patients operated in 1995 and first published in 1999. It is now 10 years old, and is based on data that are even older. Since the introduction of EuroSCORE, there has been a quantum improvement in cardiac surgical survival which occurred in the first two to three years of the new millennium. Evidence from countries with national databases suggests that mortality in some of them has approximately halved, despite gradual worsening of the risk profile of patients. In the United Kingdom, for



example, mortality has fallen to approximately 55% of logistic EuroSCORE prediction, giving a UK RAMR of around 0.55. This phenomenal improvement in cardiac surgical outcomes appears to have happened in coronary surgery, valve surgery, combined surgery and other procedures. Yet there has been no new discovery, drug or technological wizardry to explain such a large improvement. How did it come about?

In 1955, Landsberger (1958) analysed experiments from 1924 to 1932 at the Hawthorne Works (a Western Electric plant near Chicago). The company had commissioned a study to see if its workers would become more productive in stronger or weaker ambient light. Productivity improved when lighting was changed in either direction and worsened when the study was finished. He concluded that the improvement was due to the workers being motivated by the interest shown in them. When other changes were made and their effect similarly monitored, such as moving work stations, a similar improvement in productivity also resulted. The term Hawthorne effect was coined to describe the improvement that occurs due to the simple introduction of the monitoring of outcomes: in other words, when you measure performance, it improves.

Until the widespread use of EuroSCORE, there was no established measure of cardiac surgeons' clinical performance. EuroSCORE provided the tool for such measurement, and performance improved. As a result of this improvement, there is now evidence that the model is no longer appropriately calibrated, although its discrimination may still be powerful.

EuroSCORE has therefore probably fallen victim to its own success, in that the heightened awareness of the importance of clinical outcomes has resulted in improvement which may have made the model obsolete. Can we conclude from the above that EuroSCORE is no longer useful for the assessment of today's cardiac surgical outcomes? The answer may indeed be yes, but that requires some new data.

4. The question of calibration

If the improvement in outcomes has occurred across all cardiac surgery, then recalibration can be made easy. In order to answer the calibration question, we need a large, multi-institutional study. Most recently published studies which were multi-institutional (Yap et al., 2006; Jin and Grunkemeier, 2006) found overprediction to be the main problem, but there is an inherent publication bias with institutions which identify underprediction being less likely, willing or able to publish.

In the meantime, users of EuroSCORE can be assured that it is still a valuable tool for assessing cardiac surgical risk. As most studies show, any risk model offers a set standard. Some units will perform at that standard, some will do better and some worse. The best estimate for evaluating the risk of mortality for a patient undergoing a particular procedure at a particular institution is to calculate the logistic EuroSCORE (Roques et al., 2003) and then to correct it for the performance of the unit in question, so that the patient should be quoted a predicted mortality calculated by multiplying the patient's EuroSCORE by the hospital's RAMR as follows:

predicted mortality

$$= \frac{\text{patient logistic EuroSCORE} \times \text{hospital mortality}}{\text{hospital logistic EuroSCORE}}$$

In other words, in Hospital X, where the RAMR is 0.7 (actual mortality = 0.7 of predicted), the patient's logistic EuroSCORE is multiplied by 0.7, reflecting the hospital's performance. This becomes crucial if the model is being used to direct patients towards a treatment such as transcatheter aortic valve implantation (TAVI) on the basis of risk prediction. If, for example, it is decided that TAVI should be considered when the risk of AVR is 20%, it should be 20% *at the institution in question*, and not simply if the EuroSCORE is 20%. If the institution has an RAMR of 1.0, then EuroSCORE can be used. If the institution has an RAMR of 0.5, then TAVI should only be offered to patients whose EuroSCORE is 40 or above.

5. The question of discrimination

It is still reasonable to state that EuroSCORE remains powerful in discriminating between low-risk and high-risk patients: to this day the area under the ROC curve in many studies, is around 0.8 or higher. Despite this, there are probably areas for improvement. Most notable are that the model ignores the size of the intervention in many cases, has abrupt cut-off points for some continuous variables (like pulmonary hypertension, renal function, recency of myocardial infarction and others). It also ignores hepatic disease, obesity and diabetes.

6. The future: an invitation

The time has therefore come improve the model so that it can be fit for purpose in cardiac surgery of the current era. We are now embarking on a major study to update the model and to refine the risk factors and their assessment so as to improve not just calibration, but also discrimination.

We could do this in two ways. One way is to use recent data from the many databases of cardiac surgery that exist throughout the world. Another is to collect new data. Most databases are unfortunately not validated and we are not able to verify their quality. One reason for the success of EuroSCORE is that it was built on a robust and "clean" database, collected by volunteer centres. We would like to keep that standard, and will therefore collect new data.

This time, participation will be easier. In 1995, we asked for around a hundred data points on each patient. The original risk model has already discarded most of these factors as unhelpful in risk assessment. Conversely, evidence from studies suggests that new risk factors should be added and old ones refined. This time, we shall ask for only around 45 data points and for the majority of patients only about 20 data points will apply. Most data will be familiar to anyone who already uses EuroSCORE, with a few additions and refinements. These will be in the areas of diabetes, obesity, renal and liver function, the weight of the proposed intervention and one or two other fields. Data collection will be online via a dedicated website. There will be an option to collect paper-based data for later online transfer for centres requiring that facility. Data will be confidential, guaranteeing patient, centre and surgeon anonymity to all participants.

It is crucial that the data come from all types of centre, and not only so-called "centres of excellence". The entire spectrum of performance must be included and all deaths must be reported so that the risk model reflects the reality of cardiac

surgical outcomes. In fact, if only the centres with the best outcomes participate, the risk model will be harsh on everyone else.

The website is under construction and will be both robust and user-friendly. Data collection will begin early in 2010. The exact time period for collecting data is not yet fixed, because that will depend on the number of centres participating. The more centres that participate, the quicker the data collection will be completed. Already more than 200 centres have committed to the project. We hope that number will increase very rapidly. If enough centres participate, we may have data on 40 000 patients in less than 3 weeks. The longest we shall ask for will be 3 months.

Centres in Saudi Arabia and the Middle East with an interest in participating in this important initiative are invited to register their interest by following the link from the EuroSCORE website (www.euroscore.org). Once interest is registered, the team will get in touch with additional information and updates. Any potential participants who have further queries about the project can contact us by email (euroscore@papworth.nhs.uk).

We are the same scientific team that developed the original EuroSCORE. We remain independent of national and international specialist societies, governments and industry. The project is currently funded by a scientific grant from Edwards Laboratories, by Papworth Hospital, Cambridge and by the Karolinska University Hospital in Stockholm. Additional

funding sources are also being sought. Our aim is to produce the finest and most practical risk model for cardiac surgeons and their patients everywhere.

References

- Jin, R., Grunkemeier, G.L., 2006. Does the logistic EuroSCORE offer an advantage over the additive model? *Interact. Cardiovasc. Thorac. Surg.* 5, 15–17.
- Landsberger, H.A., 1958. *Hawthorne Revisited*. Cornell University Press, Ithaca, NY.
- Nashef, S.A.M., Roques, F., Michel, P., Gauducheau, E., Lemeshow, S., Salamon, R., the EuroSCORE study group, 1999. European system for cardiac operative risk evaluation (EuroSCORE). *Eur. J. Cardiothorac. Surg.* 16, 9–13.
- Roques, F., Michel, P., Goldstone, A., Nashef, S.A.M., 2003. The logistic EuroSCORE. *Eur. Heart J.* 24, 882.
- Yap, C.H., Reid, C., Yip, M., Rowland, M.A., Mohajeri, M., Skillington, P.D., Seevanayagam, S., Smith, J.A., 2006. Validation of the EuroSCORE model in Australia. *Eur. J. Cardiothorac. Surg.* 29, 441–446.

Samer A.M. Nashef

Papworth Hospital, Cambridge CB23 3RE, United Kingdom

Tel.: +44 1480 364299; fax: +44 1480 364744

E-mail address: sam.nashef@papworth.nhs.uk